

Effectiveness of Wavelet Scalogram on Partial Discharge Pattern Classification of XLPE Cable Insulation

Rakesh Sahoo^{ID} and Subrata Karmakar^{ID}, *Senior Member, IEEE*

Abstract—Detection and classification of partial discharge (PD) signal in XLPE cable is very important to find out the root cause of insulation failure. The purpose of this work is to improve the accuracy and effectiveness in classifying various types of PD signals including corona, surface, and internal discharges in a noisy environment using machine learning (ML) and deep learning (DL) techniques. For ML models, statistical features such as skewness, kurtosis, standard deviation, and variance are extracted from the PD signals and used for training. In contrast, the DL model focuses on continuous wavelet transforms (CWTs) based scalogram images which highlight the resolution of the signal's energy and are used for feature extraction in the classification study. Instead of using traditional phase-resolved PD (PRPD) plots, this work introduces a novel approach to converting PD signals to scalogram images using CWT. These scalogram images provide a visual representation of the frequency components of different PD types and how they change over time. The wavelet-based scalogram images are used as input features for the proposed convolutional neural network (CNN) and state-of-the-art (SOTA) DL models, along with shallow ML classifiers for model training, validation, and testing. Along with this, k-fold cross-validation and hyperparameter tuning are also employed to enhance the overall performance of the model. The classification results demonstrate that the proposed CNN model achieves up to 97.33% recognition accuracy and less computational complexity as compared to various ML and DL models.

Index Terms—Convolutional neural network (CNN), partial discharge (PD), scalogram image, transfer learning, wavelet transform.

I. INTRODUCTION

THE insulation property of high voltage (HV) equipment is determined by examining the characteristics and pattern of partial discharge (PD), which is an essential aspect of condition monitoring. PD occurs in the form of the corona, surface, and internal discharge types, which initiates the degradation of insulation in HV equipment [1]. For this reason, early detection and analysis of the behavior of insulation through continuous PD monitoring have a significant impact on the overall power system, which is done by analyzing the

Manuscript received 8 August 2023; revised 12 December 2023; accepted 2 January 2024. Date of publication 8 February 2024; date of current version 19 February 2024. This work was supported by the Department of Science and Technology, Government of India, through the Science and Engineering Research Board under Grant SPR/2022/000179. The Associate Editor coordinating the review process was Dr. Chao Tan. (*Corresponding author: Subrata Karmakar*)

The authors are with the National Institute of Technology Rourkela, Rourkela 769008, India (e-mail: karmakar.subrata@gmail.com).

Digital Object Identifier 10.1109/TIM.2024.3363790

magnitude, frequency, and pattern of PD [2]. For this reason, the identification and classification of PD based on its severity level has become an important area of research in recent years.

A lot of research has been carried out on the classification of PD patterns for different types of defects, including PD data acquisition, denoising, and feature selection. The goal is to improve the accuracy and reliability of PD patterns, as well as to facilitate the understanding and diagnosis of the health of insulation systems as a whole. The phase-resolved PD (PRPD) approach is one effective method for extracting characteristics from PD signals. This approach is involved in analyzing PD signals at regular intervals throughout the 360° of an ac power cycle by extracting the magnitude and number of charges [3], [4]. Examining these parameters at different phases of the power cycle is characterized by the PD activity. The effectiveness of complex and nonlinear classifiers is greatly influenced by the quality of these extracted features. To overcome this issue, researchers have suggested integrating statistical features with the PRPD method [4], [5]. These statistical features, such as skewness (indicating distribution asymmetry), kurtosis (measuring distribution sharpness), and cross correlation factor (distinguishing distribution shapes) are commonly employed. This combination aims to improve classification outcomes by providing valuable information for differentiating between various PD patterns associated with defects. Recent studies have focused not only on statistical features but also on spectral, fractal, Weibull distribution, principal component analysis (PCA), time-frequency-based scalogram, and wavelet transform features for pattern analysis [6], [7], [8]. Along with this machine learning (ML) and deep learning (DL) methods are becoming increasingly used for pattern recognition and severity classification of discharge signals. Probabilistic neural networks (PNNs), support vector machines (SVMs), random forests (RFs), extension neural networks (ENN), and artificial neural networks (ANNs) are common ML techniques used for classification problems after feature extraction from the PD signals [3], [4], [10], [11], [12]. In addition, DL methods, particularly the deep neural network (DNN) architecture, have demonstrated promising results in feature extraction and classification tasks. Unlike traditional ML approaches, DL methods eliminate manual feature engineering by leveraging multiple layers of nonlinear transformations to learn essential features such as edges, textures, shapes, and patterns at different scales directly from raw input data [13], [14], [15]. This capability makes them

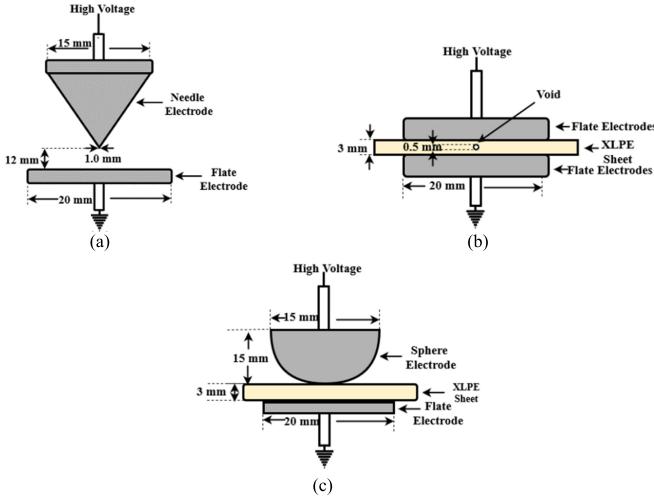


Fig. 1. Schematic representation of electrode configuration for (a) corona, (b) internal, and (c) surface PD.

highly effective in image classification. DL models, including autoencoders (AEs), recurrent neural networks (RNNs), convolutional neural networks (CNNs), generative adversarial networks (GANs), and others have been successfully employed in PD detection, classification, and identification of HV insulation degradation levels [16], [17], [18], [19], [20], [21].

In this study, a continuous wavelet transform (CWT) based time-frequency mapping is built for the detection and classification of three distinct types of PD signals using the DL technique. A total of 1852 PD signals are collected from the HV laboratory having a duration of 20 ms each for the classification study. Then, the CWT is used to analyze the signal strength at various frequencies with multiple resolutions. The resulting time-frequency mapping has shown the distribution of different PD signals' energy at different frequencies, which is essential for feature extraction in the DL model. The study used a total of 1200 clean PD signals for model training, and validation. After model building the classification task is performed on the preserved test data of 326 each, for both clean and noisy PD signals. The obtained classification performance is compared with ML, proposed CNN, and the existing state-of-the-art (SOTA)-based DL models (VGG19, ResNet50, Xception, DenseNET201, NasNetLarge, EfficientNetV2S, and ConvNeXtTiny) along with shallow ML classifier algorithms including SVM, RF, and kNN. Both k-fold cross-validation and hyperparameter tuning are also employed to enhance overall performance.

II. EXPERIMENTAL DETAILS

In this study, a variety of electrode configurations, such as needle-flat, sphere-flat, and flat-flat, are utilized to generate three distinct categories of PD signals including corona, surface, and internal which are depicted in Fig. 1. The samples are sliced from a 3 core, HV 33/33 kV unearthed XLPE insulated cable, having the cross-sectional area of 150 mm².

Fig. 1(a) depicts a needle-flat electrode arrangement that allows corona discharge to develop, with a needle-tip radius of 1 mm and a gap distance of 12 mm. For the internal PD



Fig. 2. Photograph of the experimental setup, 1. HV transformer, 2. Coupling Capacitor (C_k), 3. Measuring Capacitor (C_m), 4. Test object, 5. Needle-flat type, 6. Flat-flat type, and 7. Sphere-flat type electrode arrangement.

as shown in Fig. 1(b), a flat-flat electrode configuration is used, and an XLPE insulation sheet consisting of 0.5 mm void is placed. To produce a surface discharge type, as shown in Fig. 1(c), a 3 mm wide XLPE sheet is sandwiched between sphere-flat electrode arrangements. All of these electrode configurations and setups are immersed in insulating oil for the generation and observation of PD signals.

The experimental setup as per IEC 60270 is placed in an HV laboratory, which is a noise-free chamber with a controlled temperature of $24^\circ\text{C} \pm 2^\circ\text{C}$. The experimental setup shown in Fig. 2 consists of a high-voltage transformer, a coupling capacitor (C_k), a measuring capacitor (C_m), a test cell, and a PD detection system [14]. The PD detection system consists of an AKV-DU coupling quadripole and a DTM PD meter. The AKV-DU coupling quadripole is used to couple the PD signals from the test cell to the PD meter. The DTM PD meter measures the PD intensity and displays it on a screen. The PD signals are then stored in a personal computer for further analysis.

III. WAVELET-BASED SCALOGRAM ANALYSIS

The spectrogram, based on the Fourier Transform, is commonly employed for the analysis and classification of various PD sources. However, in some cases, the spectrogram may not be sufficient to accurately represent the nonstationary PD signals. As an alternative, the CWT-based scalogram is used as a replacement representation. In contrast to the short-time Fourier transform (STFT) spectrogram, the CWT offers improved time-frequency resolution, making it particularly suitable for analyzing signals with rapidly changing frequency content, such as PD signals. This is achieved by convolving the signal with a family of wavelet functions at different scales, thereby capturing both localized and global frequency components

$$\text{STFT}\{x(t)\}(\tau, \omega) = \int_{-\infty}^{\infty} x(t)h(t - \tau)e^{-j\omega t} dt. \quad (1)$$

The STFT is a mathematical operation that converts a 1-D signal into a 2-D function in the time-frequency plane as shown in (1). It involves convolving the signal $x(t)$ with a

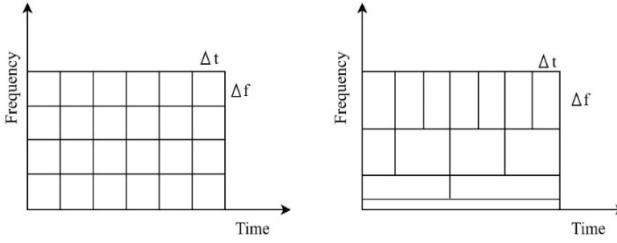


Fig. 3. Time-frequency resolution for (a) spectrogram and (b) scalogram.

small window function $h(t)$ centered at a specific time location τ . The main concern with the STFT method is the issue of resolution, i.e., only possible to determine the time intervals during which specific frequency bands exist, rather than being able to identify all of the frequency components present at any given time. This resolution problem has limited the ability of the STFT to capture rapidly changing frequency components in a signal. For this reason, to overcome this the CWT is used by providing improved time-frequency resolution [22]

$$\text{CWT}\{x(t)\} = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)g\left(\frac{t-\tau}{a}\right)dt. \quad (2)$$

Equation (2) is expressed as the CWT mathematical representation where a and τ represent the scaling and transformation parameters, respectively. The mother wavelet is represented as $g(t, \tau)$, and normalization of the energy content is achieved using $(1/(a)^{1/2})$. The CWT analysis results in a representation known as a scalogram, which depicts the distribution of the signal's energy content in the time-scale domain. Valuable insights into the time-frequency characteristics of the signal are provided by the scalogram. In Fig. 3, a comparison of time-frequency resolutions between STFT-based spectrograms and CWT-based scalograms is shown. Fig. 3(a) depicts a spectrogram based on the STFT, which encounters a resolution issue because of an inequality constraint. The spectrogram divides the window length into equal-sized sections, leading to limited time-frequency resolutions. In contrast to the STFT-based spectrogram, the CWT divides the window length for a specific area into different proportions with respect to time and frequency. This distribution of window lengths is shown in Fig. 3(b). The CWT provides different time-frequency resolutions compared to the spectrogram. This behavior is similar to that of PD signals, which exhibit varying characteristics in both time and frequency domains.

IV. ARCHITECTURES OF PD CLASSIFIER

The development of intelligent solutions to complex problems such as pattern analysis is enabled by ML and DL architecture. In this section, the architectures of the ML and DL model along with proposed CNN and SOTA-based transfer learning approaches are used to classify three distinct kinds of discharge signals, as shown in Fig. 4.

In the process of ML model building, statistical features such as skewness, kurtosis, standard deviation, and variance are extracted from the PD signals. These features are then utilized to train classifiers such as ANN, SVM, RF, and kNN. CNNs are designed to automatically learn and extract features

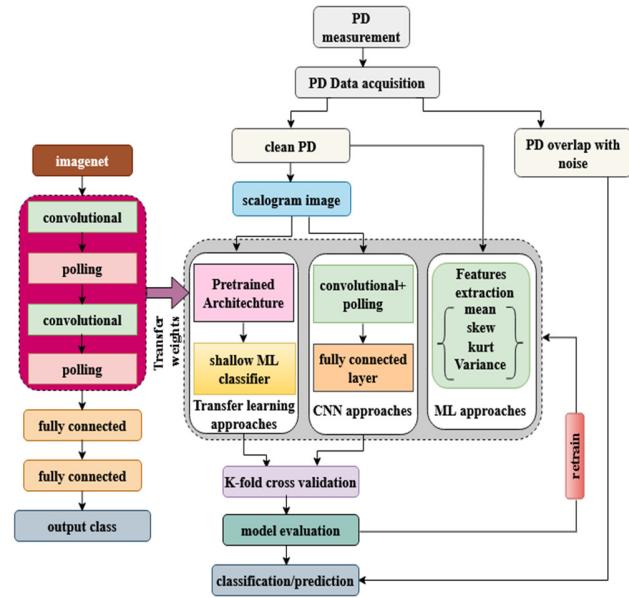


Fig. 4. Model architecture for classification study.

from image data, by using a combination of convolutional layers, pooling layers, and fully connected layers. The first few layers of a CNN are usually designed to perform feature learning, where the network learns to detect and extract low-level features from the input data. These layers typically consist of convolutional layers, which use small filters that move across the input image, performing convolutions to extract local features such as edges, corners, and textures. The pooling layers then downsample the feature maps to reduce their spatial size, while preserving the most important features. Finally, the classification or regression tasks are performed using the fully connected layers based on the extracted features. In this work, the architecture of the pre-trained NN model consists of fixed feature learning along with shallow ML-based classification subnetworks.

To implement transfer learning, a pre-trained NN's final fully connected layer is replaced by a custom neural network, and the network is retrained on a new dataset. Typically, the pre-trained NN has been trained on a large dataset, such as ImageNet, to perform a specific task, such as image classification. While using transfer learning, the weights of the convolutional layers and pooling layers of the pre-trained NN are frozen, and only the weights of the fully connected layers are modified during training on the new task. The final fully connected layer of the pre-trained network is replaced with a new custom neural network that is suitable for the new task. This study focused on seven distinct varieties of SOTA-based DL model architecture, which include ResNet50, VGG19, Xception, DenseNet201, NasNetLarge, EfficientNetV2S, and ConvNeXtTiny [23].

For the DL model, scalogram images are formed in RGB scale, and automated features are extracted which is used for model building. The final model performance is evaluated on the preserved noisy test-set scalogram images and a classification score is obtained. In both, the case after

TABLE I
HYPERPARAMETERS OF ML AND DL MODEL

Classifier	Hyperparameters	Type and set range
CNN	Optimizer	Adagrad, Adamax, RMSprop, Adam
	Learning rate	0.0001-0.001
	Activation function	ReLU, ELU, Leaky ReLU
	Kernel size	3x3, 5x5, 7x7
ANN	Pooling type	Max pooling, Average polling
	Optimizer	Adagrad, Adamax, Adam
	Activation function	ReLU, Sigmoid
	Batch size	16, 32
SVM	kernel	Gaussian, RBF
	Gamma (γ)	$1 < \gamma < 0.0001$
RF	Regularization parameter (C)	$0.1 < C < 1000$
	n_estimators	$10 < n_estimators < 200$
KNN	Number of neighbours (K)	$1 < K < 20$

TABLE II
COMPARISON OF INCEPTION AND EXTINCTION VOLTAGE WITH VARIOUS ELECTRODE ARRANGEMENTS

Sr. No.	PD Type	Electrode arrangement	PDIV (kV)	PDEV (kV)
1	Corona	Needle-flat	8.2	7.3
2	Internal	Flat-flat	7.6	7.1
3	Surface	Sphere-flat	8.5	7.7

fivefold cross-validation best parameters are analyzed in terms of accuracy. If the obtained result is not up to the mark, then retraining is done with the help of the selection of hyperparameters. In this study, the tuning of hyperparameters is done using the grid search technique shown in Table I [24], [25].

V. RESULTS AND DISCUSSION

A. Experimental Outcome and Scalogram-Based Pattern Analysis

The experiment is designed to collect different types of PD signals that are applicable for classification purposes using ML and DL approaches. The collected data is used to train algorithms, enabling them to classify various types of PD signals, which could have applications in the detection and diagnosis of electrical insulation problems. Table II provides information about the PD signals obtained through various electrode configurations, comprising the PD Inception Voltage (PDIV) and PD Extinction Voltage (PDEV) for each signal kind. Wavelet scalogram is a visualization technique that mainly emphasizes nonstationary PD signal which shows the frequency content that changes over time. This information is used to identify patterns or anomalies in the signal, which helps us to detect and diagnose the insulation. The obtained scalogram images of three different categories of PD using the Morlet mother wavelet of CWT are shown in Fig. 5, where the x- and y-axis represent time and frequency, respectively [26], [27].

The color or shading of each point in the graph indicates the energy or strength of the signal at a particular time and frequency. From the obtained scalogram images based on pixel

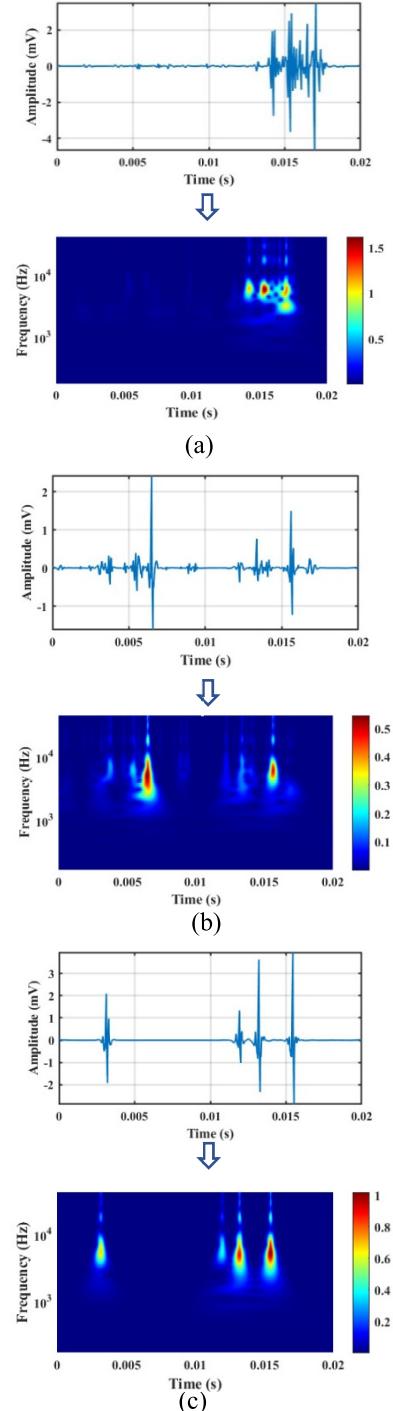


Fig. 5. Time domain and scalogram representation of (a) corona, (b) internal, and (c) surface PD.

intensity, it is quite difficult to distinguish different types of PD. For this reason, insulation defects. It also provides a visual representation of the energy distribution of the signal over different frequencies and time intervals. CNN-based DL framework is employed in this work, which can automatically extract the features and perform the classification task without the need for manual feature engineering. With regard to normalization, the Min-Max scaling approach is used to transform the raw data (scalogram image pixels). The automatic feature

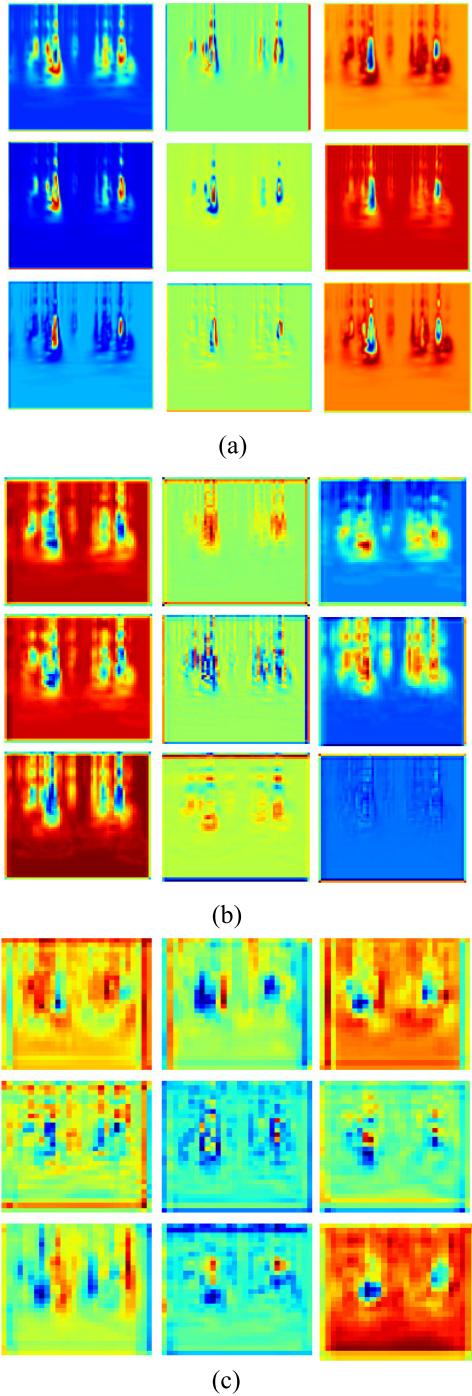


Fig. 6. Feature map of (a) conv_2d_1, (b) conv_2d_2, and (c) conv_2d_3 scalogram image.

extraction process is observed by examining the output of the filter banks. By comparing the filters in the first layer with those in the subsequent layer, Fig. 6 illustrates a transition in the feature maps. As the convolution number increases, the feature maps evolve from capturing fine details to revealing broader aspects and providing compression. In this work, out of 1852 collected various PD signals, 1200 of them were used for model training and validation.

The classification task is performed on the preserved test data of both clean and noisy PD signals of 326 each. The

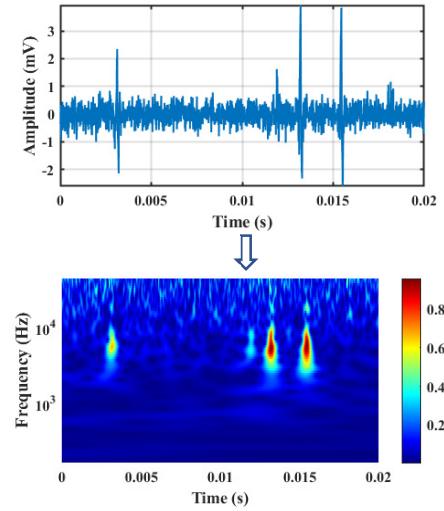


Fig. 7. One-dimensional time domain and 2-D scalogram representation of PD signal along with noise.

noisy PD signals are created by overlaying clean PD signals with noise at a signal-to-noise ratio of -15 dB, as illustrated in Fig. 7. This type of testing is important to ensure that the model is robust and accurately diagnose insulation defects even when the signals are noisy or distorted. Along with this, to enhance the model's ability during the model-building process, image augmentation techniques, including rotation, horizontal flipping, brightness adjustment, and cropping, are employed to expand the number of CWT images used for training. As a result, the original set of 1200 experimental training images is being transformed into 5690 augmented images during the DL model construction phase.

B. Classification Study

1) ML Application: In this study, for the ML model building statistical input features (mean, variance, skewness, standard deviation, and kurtosis) are extracted from the collected PD signal. The ML model mainly emphasizes ANN, SVM, kNN, and RF. The ANN model employed a supervised learning approach and initialization of random weight. The architecture utilizes three hidden layers with 8, 16, and 16 neurons, respectively, employing Adam optimization for training, ReLU activation for the hidden layers, and Softmax activation for the output layer. In SVM a hyperplane is the decision boundary line that provides a sufficient amount of separation between the classes. In addition, hyperparameters, such as kernel, gamma, and regularization values, are considered to improve the model's overall performance. In this study, the Gaussian RBF kernel having the gamma and regularization parameter (C) in the range of $1 < \text{gamma} < 0.0001$ and $0.1 < C < 1000$, respectively, is used. An RF is defined by the involvement of a large number of trees, where the forest is formed by randomly distributed trees on Earth. It is composed of several decision trees which are decided by $n_{\text{estimators}}$. This decision tree is then trained using the bagging technique. Bagging divides a feature dataset into several subgroups with replacement; the number of subgroups is determined by the number of decision trees. In this work, $n_{\text{estimators}}$ in the range of

TABLE III
MODEL ARCHITECTURE OF PROPOSED CNN

Layers	Parameters
Conv_2d_1	Filter no.: 32, kernel_size: 5×5, stride:1, padding: same
Conv_2d_2	Filter no.: 64, kernel_size: 5×5, stride:1, padding: same
Pooling_1	kernel_size: 2×2, stride:2
Conv_2d_3	Filter no.: 128, kernel_size: 5×5, stride:1, padding: same
Conv_2d_4	Filter no.: 256, kernel_size: 5×5, stride: 1, padding: same
Pooling_2	kernel_size: 2×2, stride: 2
FC_1	Feature_map_Size: 512
FC_2	Feature_map_Size: 64
Softmax	Feature_map_Size: 3

TABLE IV
ACCURACY OF DIFFERENT OPTIMIZERS WITH LEARNING RATE

lr	Adagrad	Adamax	RMSprop	Adam
0.0001	0.6565	0.9035	0.9398	0.9297
0.0002	0.7676	0.9191	0.9097	0.9297
0.0003	0.8353	0.9191	0.9097	0.9595
0.0004	0.8789	0.9292	0.9298	0.9595
0.0005	0.9393	0.9090	0.9198	0.9797
0.0006	0.8787	0.9498	0.9298	0.9494
0.0007	0.9090	0.9297	0.9096	0.9393
0.0008	0.9191	0.9296	0.9198	0.9494
0.0009	0.9232	0.9136	0.9398	0.9595
0.001	0.9292	0.9266	0.9498	0.9595

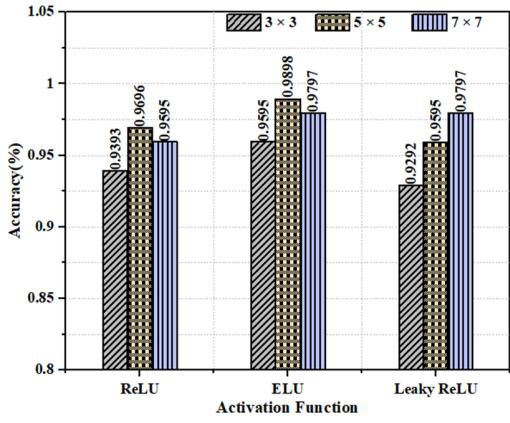


Fig. 8. Performance of activation function along with filter size.

10 < n_estimators < 200 are used. The K -nearest neighbors algorithm is the most basic classification algorithm with highly competitive results. In this work, the “ K ” value is chosen based on the error rate, with a selected range of “ K .” Here, the range of K varies from $1 < K < 20$, and the obtained error rate is minimum [0.06:K:0.14].

2) *DNN Application:* The wavelet-based scalogram images are used as input features for both the proposed CNN and SOTA DL models along with shallow ML classifiers for model training, validation, and testing.

A detailed description of the proposed CNN model architecture for a classification task is shown in Table III. Each convolution layer (conv_2d) is applied with an appropriate number of filters having an appropriate filter size to the input

TABLE V
EFFECTIVENESS OF VARIOUS POOLING METHODS

Pooling method	Accuracy
Max pooling	0.9732
Average pooling	0.9596

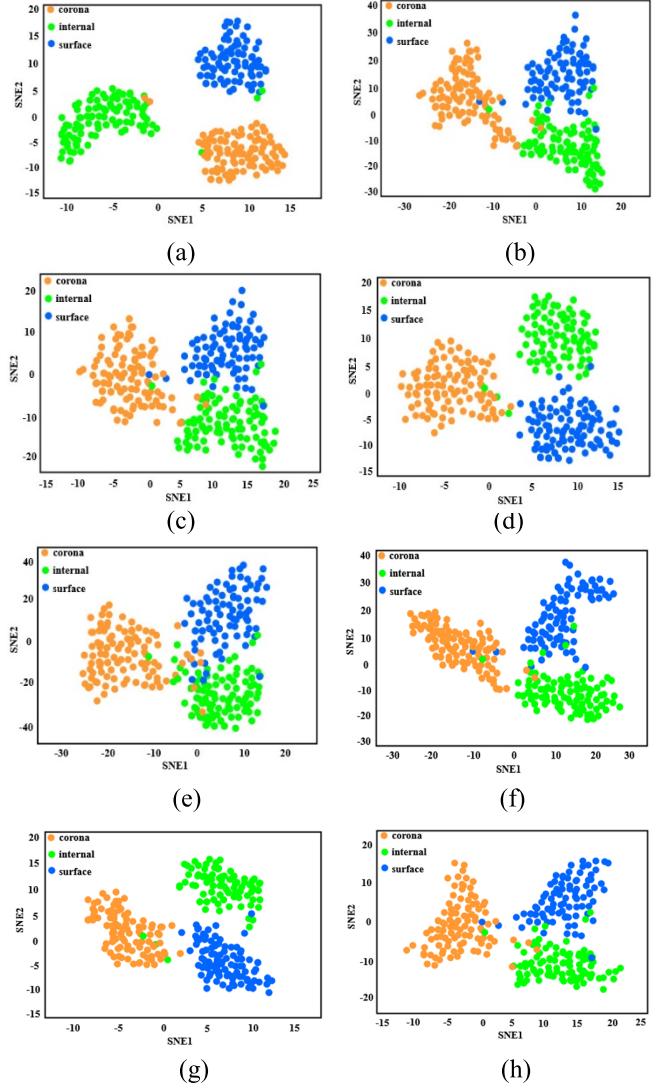


Fig. 9. t-SNE representation of (a) CNN, (b) ResNet50, (c) VGG19, (d) DenseNET201, (e) Xception, (f) NasNetLarge, (g) EfficientNetV2S, and (h) ConvNeXtTiny model.

scalogram image for extracting the input features and the dimension of the feature map is lowered by the pooling layer. Batch normalization is used after each convolutional layer. The model also included two fully connected layers (FC), which are used to perform the final classification task. The SoftMax activation function is used at the end of the model, which outputs a probability distribution over the possible classes. The dropout layer is used to reduce overfitting, and the padding is set to “same” to ensure that the model can learn features effectively. Selecting the appropriate hyperparameters for a CNN is crucial for achieving optimal performance on a given task. In this study, the accuracy of the model is analyzed by

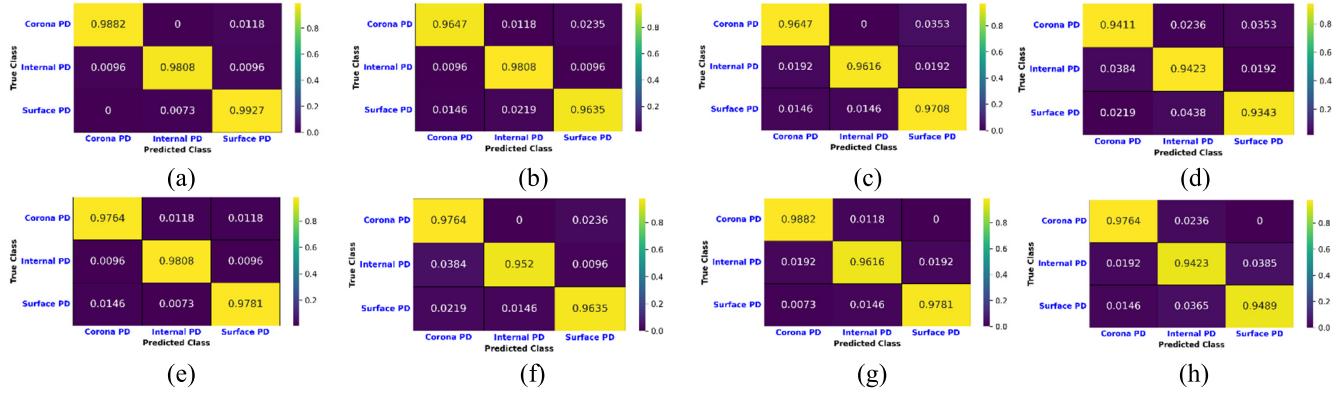


Fig. 10. Confusion matrix of (a) CNN, (b) VGG19, (c) ResNet50, (d) Xception, (e) DenseNET201, (f) NasNetLarge, (g) EfficientNetV2S, and (h) ConvNeXtTiny model in clean PD dataset.

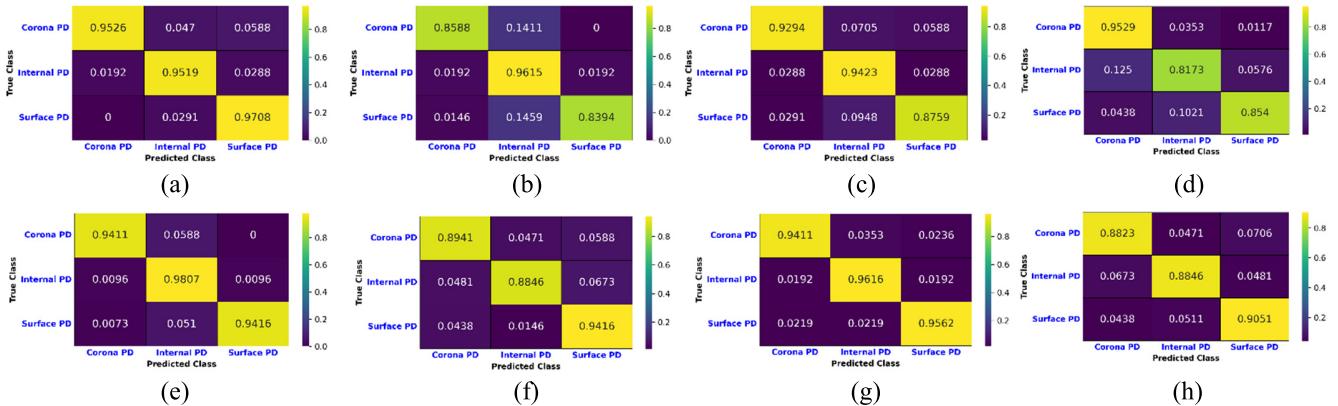


Fig. 11. Confusion matrix of (a) CNN, (b) VGG19, (c) ResNet50, (d) Xception, (e) DenseNET201, (f) NasNetLarge, (g) EfficientNetV2S, and (h) ConvNeXtTiny model in noisy PD dataset.

TABLE VI
OVERVIEW OF DIFFERENT SOTA DL MODELS

Model	Configuration	Year	Input size	Parameters
VGG	VGG19	2015	224×224	143.7
ResNet	ResNet50	2015	224×224	25.6
Xception	Xception	2017	299×299	22.9
DenseNet	DenseNet201	2017	224×224	20.2
NasNet	NasNetLarge	2018	331×331	88.9
EfficientNet	EfficientNetV2S	2021	299×299	21.6
ConvNeXt	ConvNeXtTiny	2022	224×224	28.6

TABLE VII
PERFORMANCE INDICES OF ML MODEL

ML model configuration	Performance Indices (%)					
	Clean PD			Noisy PD		
	Acc.	Sen.	Spe.	Acc.	Sen.	Spe.
ANN	93.65	90.27	95.17	87.11	79.21	90.39
SVM	94.88	92.18	96.34	88.95	82.44	91.83
RF	93.24	89.33	94.93	86.71	79.01	90.12
kNN	91.81	87.41	93.82	86.29	78.24	89.65

considering the fine-tuning of hyperparameters such as the optimizer, learning rate, activation function, filter size, and pooling methods using grid search.

The initial step involves the implementation of Adagrad, Adamax, RMSprop, and Adam optimizers to modify the network weights and decrease the losses which are presented

in Table IV. The learning rate (lr) for these optimizers is set to a value between 0.0001 and 0.001. It has been observed that the Adam optimizer achieves the maximum accuracy in comparison to other optimizers while the learning rate is set to 0.0005. After that in this classification study, the three most common types of activation functions ReLU, exponential linear unit (ELU), and Leaky ReLU are used along with the variation of kernel size 3×3 , 5×5 , 7×7 . The accuracy performance of each activation function is replicated in Fig. 8. ELU achieves the best accuracy of 100% by varying the filter kernel size to 5×5 .

The next step is to determine which form of pooling is the most effective. In this case study, Max pooling and Average pooling, with kernel size 2×2 and the stride of 2, since bigger pooling sizes might result in information losses. The accuracy in both cases is shown in Table V, which indicates the method of pooling also affects the model performance.

After performing optimal hyperparameter tuning, the ELU activation function, with 5×5 Kernel filter size and Max-pooling method, including Adam optimizer having the learning rate of 0.0005, is found to be suitable for this study.

SOTA DL models in image classification are the cutting-edge and highest-performing neural network architectures, typically employing transfer learning techniques. In this, a pre-trained DNN is used as a feature extractor, and its learned features are then used as input to a shallow ML

TABLE VIII
PERFORMANCE INDICES OF CNN AND SOTA-BASED DL MODEL WITH DIFFERENT ML CLASSIFIERS

DL model configuration	Classifier	Performance Indices (%)						Elapsed time (sec.)
		Clean PD			Noisy PD			
		Acc.	Sen.	Spe.	Acc.	Sen.	Spe.	
CNN		99.17	98.72	99.35	97.33	95.85	97.99	245.16
VGG19	SVM	97.95	96.96	98.45	90.79	87.38	93.30	442.16
	RF	97.74	96.71	98.35	92.22	88.65	94.28	438.35
	kNN	97.34	96.08	97.81	89.97	83.96	92.58	437.07
	SVM	97.74	96.63	98.31	94.08	91.58	95.64	432.17
ResNet50	RF	97.75	96.56	98.26	88.54	84.17	91.69	427.07
	kNN	97.54	96.47	98.13	89.36	84.71	92.09	425.46
	SVM	95.90	93.92	96.94	91.20	87.47	93.58	341.32
	RF	95.69	93.85	96.81	90.58	86.23	93.1	337.23
Xception	kNN	95.29	93.26	96.55	88.54	82.59	91.5	337.03
	SVM	98.36	97.59	98.74	96.92	95.44	97.74	521.48
	RF	98.56	97.84	98.92	92.43	88.73	94.47	517.05
	kNN	98.15	97.42	98.64	89.15	83.93	92.51	516.48
DenseNet201	SVM	97.74	95.28	98.88	94.06	90.67	95.45	353.09
	RF	96.33	94.19	97.19	93.24	89.39	94.87	352.56
	kNN	95.90	93.62	96.90	92.38	88.01	93.96	351.27
	SVM	98.15	97.20	98.60	95.69	93.53	96.84	327.34
NasNetLarge	RF	98.36	97.59	98.77	96.93	95.62	97.88	329.42
	kNN	97.94	96.96	98.45	95.48	92.95	95.56	326.49
	SVM	96.58	94.80	97.41	92.02	88.32	94.71	623.32
	RF	96.92	95.58	97.68	92.83	89.06	94.61	627.09
EfficientNetV2S	kNN	96.51	94.73	97.42	90.58	85.09	92.59	623.03

model. In this work, the SOTA model (ResNet50, VGG19, Xception, DenseNet201, NasNetLarge, EfficientNetV2S, and ConvNeXtTiny) is trained on the ImageNet dataset and transferred its weight to the new set of training images which is shown in Table VI. From the new set of training images, high-level features are extracted which act as input to shallow ML classifiers, such as SVM, RF, and kNN, to perform the final classification task. In addition, using a pre-trained NN as a feature extractor can reduce the amount of data needed for training the final model and also decrease the training time and computational resources required.

C. Performance Analysis

After ML and DL model building along with fivefold cross-validation, the classification task is performed on the preserved test data of both clean and noisy PD signals. Here, the preserved PD data are extracted at a different voltage level, i.e., for corona, internal, and surface PD the applied voltage is 14, 12, and 10 kV, respectively.

The performance of the model, including the sensitivity (Sen.), accuracy (Acc.), and specificity (Spe.) is assessed by evaluating the confusion matrix obtained from the test dataset. The performance indices of various ML model configurations based on statistical parameters of different PD data are displayed in Table VII.

The classification task is performed efficiently on a workstation with a high-end GPU (24 GB) and ample RAM (16 GB), enabling DL models to be trained effectively. The use of Anaconda and Jupyter Notebook (version 3.8.5) allows for easy package and dependency management, as well as visualizing the training process. The full potential of the GPU for accelerated training of DL models is enabled by the CUDA (11.5) and cuDNN (8.3.1) libraries. Visual Studio (2022) provides a robust framework for parallel computing,

which will further enhance the performance of your training process.

Fig. 9 shows the t-SNE features visualization which is obtained from CNN and pre-trained NN models.

In each model, the feature clusters are easily identified, whereas in the Xception model, the cluster boundaries are allied together as compared to the proposed CNN. Figs. 10 and 11 illustrate the confusion matrix of the proposed CNN along with the SOTA-based DL model. It contains the details of testing patterns of various models on the proportion of accurate and inaccurate predictions [28]. A comparative study of the proposed CNN along with pre-trained architecture is shown in Table VIII, which classifies different PD signals in terms of performance indices. This shows that the proposed CNN model based on DL is better at classifying different PD events.

It is demonstrated that CNN is superior to other pre-trained models including ML classifiers in the categorization of corona PD, internal PD, and surface PD. Many researchers have classified different PD types on the basis of PD pattern analysis using an ML algorithm based on manual feature attribute selection. However, the CWT-based scalogram picture is highlighted in this work and is used to choose features without requiring manual feature engineering for the classification study.

VI. CONCLUSION

The primary objective of this work is to classify three distinct categories of PD signals in the presence of noise. These signals are experimentally collected from the HV laboratory and used a total of 1200 clean PD signals for training, and validation. After ML and DL model building the classification task is performed on the preserved test data of both clean and noisy PD signals of 326 each. To classify PD

signals, for ML models, statistical features such as skewness, kurtosis, standard deviation, and variance are extracted from the signals and used for training. In contrast, the DL model focuses on CWT-based scalogram images which highlight the resolution of the signal's energy and are used for feature extraction in the classification study. The collected signal is converted to CWT-based scalogram images which highlight the frequency components of different PD types and how they change with respect to time instead of traditional PRPD. In this work, DL methods including CNN and SOTA-based DL models provide the advantage of automatically learning relevant features from the CWT coefficients, allowing for end-to-end feature extraction and classification. The resolution of different PD signals' energy is distributed differently in the scalogram images, which is vital for extracting features in proposed CNN and SOTA-based DL models (VGG19, ResNet50, Xception, DenseNET201, NasNetLarge, EfficientNetV2S, and ConvNeXtTiny). In this framework, the process of extracting features is automated, which is very advanced and gets rid of the need for human intervention. Even if DL's feature selection process is automated, a hyperparameter setting is still required for optimal classification performance which is analyzed in this work. The classification result is also compared with the various pre-trained models including shallow ML classifier (SVM, RF, and kNN). The proposed CNN with optimal hyperparameter tuning has achieved 99.18% and 97.33% classification accuracy in clean PD and noisy PD signal datasets, respectively, as compared to both the ML and SOTA-based DL models with transfer learning approaches. In recognition of accuracy and speed, the choice of network types and feature settings has a significant impact on categorizing different PD types and preventing failure of insulation. In the future PD signal classification can be done by exploring different feature extraction techniques, incorporating data augmentation, integrating multiple data sources, testing in practical applications, and improving model interpretability.

REFERENCES

- [1] S. Lu, H. Chai, A. Sahoo, and B. T. Phung, "Condition monitoring based on partial discharge diagnostics using machine learning methods: A comprehensive state-of-the-art review," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 27, no. 6, pp. 1861–1888, Dec. 2020, doi: [10.1109/TDEI.2020.009070](https://doi.org/10.1109/TDEI.2020.009070).
- [2] Y. Han and Y. H. Song, "Condition monitoring techniques for electrical equipment—A literature survey," *IEEE Trans. Power Del.*, vol. 18, no. 1, pp. 4–13, Jan. 2003, doi: [10.1109/TPWRD.2002.801425](https://doi.org/10.1109/TPWRD.2002.801425).
- [3] L. Hao and P. Lewin, "Partial discharge source discrimination using a support vector machine," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 1, pp. 189–197, Feb. 2010, doi: [10.1109/TDEI.2010.5412017](https://doi.org/10.1109/TDEI.2010.5412017).
- [4] M. Karimi, M. Majidi, H. MirSaeedi, M. M. Arefi, and M. Oskouee, "A novel application of deep belief networks in learning partial discharge patterns for classifying corona, surface, and internal discharges," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3277–3287, Apr. 2020, doi: [10.1109/TIE.2019.2908580](https://doi.org/10.1109/TIE.2019.2908580).
- [5] O. Perpiñán, M. A. Sánchez-Urán, F. Álvarez, J. Ortego, and F. Garnacho, "Signal analysis and feature generation for pattern identification of partial discharges in high-voltage equipment," *Electr. Power Syst. Res.*, vol. 95, pp. 56–65, Feb. 2013, doi: [10.1016/j.epsr.2012.08.016](https://doi.org/10.1016/j.epsr.2012.08.016).
- [6] J. A. Ardila-Rey, R. Schurch, N. M. Poblete, S. Govindarajan, O. Muñoz, and B. A. de Castro, "Separation of partial discharges sources and noise based on the temporal and spectral response of the signals," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021, doi: [10.1109/TIM.2021.3121488](https://doi.org/10.1109/TIM.2021.3121488).
- [7] M. Ghorat, G. B. Gharehpétian, H. Latifi, and M. A. Hejazi, "A new partial discharge signal denoising algorithm based on adaptive dual-tree complex wavelet transform," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 10, pp. 2262–2272, Oct. 2018, doi: [10.1109/TIM.2018.2816438](https://doi.org/10.1109/TIM.2018.2816438).
- [8] C. Xia et al., "Multispectral optical partial discharge detection, recognition, and assessment," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022, doi: [10.1109/TIM.2022.3162284](https://doi.org/10.1109/TIM.2022.3162284).
- [9] A. A. Soltani and A. El-Hag, "A new radial basis function neural network-based method for denoising of partial discharge signals," *Measurement*, vol. 172, Feb. 2021, Art. no. 108970, doi: [10.1016/j.measurement.2021.108970](https://doi.org/10.1016/j.measurement.2021.108970).
- [10] A. A. Mas'ud, B. G. Stewart, and S. G. McMeekin, "Application of an ensemble neural network for classifying partial discharge patterns," *Electr. Power Syst. Res.*, vol. 110, pp. 154–162, May 2014, doi: [10.1016/j.epsr.2014.01.010](https://doi.org/10.1016/j.epsr.2014.01.010).
- [11] H. Janani, S. Shahabi, and B. Kordi, "Separation and classification of concurrent partial discharge signals using statistical-based feature analysis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 27, no. 6, pp. 1933–1941, Dec. 2020.
- [12] Q. Jing, J. Yan, Y. Wang, R. He, and L. Lu, "A novel differentiable neural network architecture automatic search method for GIS partial discharge pattern recognition," *Measurement*, vol. 195, May 2022, Art. no. 111154, doi: [10.1016/j.measurement.2022.111154](https://doi.org/10.1016/j.measurement.2022.111154).
- [13] S. Barrios, D. Buldain, M. P. Comech, I. Gilbert, and I. Orue, "Partial discharge classification using deep learning methods—Survey of recent progress," *Energies*, vol. 12, no. 13, p. 2485, Jun. 2019, doi: [10.3390/en12132485](https://doi.org/10.3390/en12132485).
- [14] R. Sahoo and S. Karmakar, "Investigation of electrical tree growth characteristics and partial discharge pattern analysis using deep neural network," *Electr. Power Syst. Res.*, vol. 220, Jul. 2023, Art. no. 109287, doi: [10.1016/j.epsr.2023.109287](https://doi.org/10.1016/j.epsr.2023.109287).
- [15] M. Florkowski, "Classification of partial discharge images using deep convolutional neural networks," *Energies*, vol. 13, no. 20, p. 5496, Oct. 2020, doi: [10.3390/en13205496](https://doi.org/10.3390/en13205496).
- [16] F. Yuwei et al., "Partial discharge pattern recognition method based on transfer learning and DenseNet model," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 30, no. 3, pp. 1240–1246, Jun. 2023, doi: [10.1109/TDEI.2023.3239032](https://doi.org/10.1109/TDEI.2023.3239032).
- [17] O. Aldosari, M. A. Aldowsari, S. M. Batiyah, and N. Kanagaraj, "Image-based partial discharge identification in high voltage cables using hybrid deep network," *IEEE Access*, vol. 11, pp. 50325–50333, 2023, doi: [10.1109/ACCESS.2023.3278054](https://doi.org/10.1109/ACCESS.2023.3278054).
- [18] Q. Zheng et al., "A real-time transformer discharge pattern recognition method based on CNN-LSTM driven by few-shot learning," *Electr. Power Syst. Res.*, vol. 219, Jun. 2023, Art. no. 109241, doi: [10.1016/j.epsr.2023.109241](https://doi.org/10.1016/j.epsr.2023.109241).
- [19] S. Sun et al., "Fault diagnosis of conventional circuit breaker contact system based on time-frequency analysis and improved AlexNet," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021, doi: [10.1109/TIM.2020.3045798](https://doi.org/10.1109/TIM.2020.3045798).
- [20] G. Zhu, K. Zhou, L. Lu, Y. Fu, Z. Liu, and X. Yang, "Partial discharge data augmentation based on improved Wasserstein generative adversarial network with gradient penalty," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6565–6575, May 2023, doi: [10.1109/TII.2022.3197839](https://doi.org/10.1109/TII.2022.3197839).
- [21] M.-T. Nguyen, V.-H. Nguyen, S.-J. Yun, and Y.-H. Kim, "Recurrent neural network for partial discharge diagnosis in gas-insulated switchgear," *Energies*, vol. 11, no. 5, p. 1202, May 2018, doi: [10.3390/en11051202](https://doi.org/10.3390/en11051202).
- [22] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990, doi: [10.1109/18.57199](https://doi.org/10.1109/18.57199).
- [23] A. Géron, *Hands-On Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA, USA: O'Reilly, 2019.
- [24] T.-D. Do, V.-N. Tuyet-Doan, Y.-S. Cho, J.-H. Sun, and Y.-H. Kim, "Convolutional-neural-network-based partial discharge diagnosis for power transformer using UHF sensor," *IEEE Access*, vol. 8, pp. 207377–207388, 2020.

- [25] A. K. Das, D. Dey, B. Chatterjee, and S. Dalai, "A transfer learning approach to sense the degree of surface pollution for metal oxide surge arrester employing infrared thermal imaging," *IEEE Sensors J.*, vol. 21, no. 15, pp. 16961–16968, Aug. 2021, doi: [10.1109/JSEN.2021.3079570](https://doi.org/10.1109/JSEN.2021.3079570).
- [26] S. Barrios, D. Buldain, M. P. Comech, and I. Gilbert, "Partial discharge identification in MV switchgear using scalogram representations and convolutional AutoEncoder," *IEEE Trans. Power Del.*, vol. 36, no. 6, pp. 3448–3455, Dec. 2021, doi: [10.1109/TPWRD.2020.3042934](https://doi.org/10.1109/TPWRD.2020.3042934).
- [27] H. I. Uckol and S. Ilhan, "Identification of corona discharges based on wavelet scalogram images with deep convolutional neural networks," *Electr. Power Syst. Res.*, vol. 224, Nov. 2023, Art. no. 109712, doi: [10.1016/j.epsr.2023.109712](https://doi.org/10.1016/j.epsr.2023.109712).
- [28] K. Al-Jabery, T. Obafemi-Ajayi, G. Olbricht, and D. Wunsch, *Computational Learning Approaches to Data Analytics in Biomedical Applications*. New York, NY, USA: Academic, 2019, doi: [10.1016/C2016-0-04633-8](https://doi.org/10.1016/C2016-0-04633-8).



Subrata Karmakar (Senior Member, IEEE) was born in 1981. He received the bachelor's degree (Hons.) in electrical engineering from Burdwan University, Bardhaman, India, in 2004, and the M.Tech. and Ph.D. degrees from the National Institute of Technology, Durgapur, India in 2006 and 2011, respectively.

He is currently an Associate Professor with the Department of Electrical Engineering, National Institute of Technology Rourkela (NIT) Rourkela, Rourkela, India. His current research interests include condition monitoring of high voltage equipment, dielectrics fluids, detection and localization of faults using signal processing, machine learning, and deep learning technique in high voltage power equipment.

Dr. Karmakar is currently a member of IEEE Dielectric and Electrical Insulation Society (DEIS).



Rakesh Sahoo was born in India, in 1990. He received the B.Tech. degree in electrical and electronics engineering from the Biju Patnaik University of Technology, Rourkela, India, in 2011, and the M.Tech. degree in electrical engineering from the Veer Surendra Sai University of Technology, Burla, Sambalpur, Odisha, India, in 2014. He is currently pursuing the Ph.D. degree in electrical engineering from the National Institute of Technology Rourkela, Rourkela.

His current research interests include pattern recognition, electrical and physicochemical characteristics analysis as well as condition monitoring of high-voltage underground cables using machine learning, and deep learning methods.